https://doi.org/10.65025/ICAIC25193m

Enhanced Evaluation System for Student Performance Analysis

Ashwini Matange
B.Tech, Computer Science Engineering
Pimpri Chinchwad College of Engineering, Pune, India
ashwini.matange@pccoepune.org

Akshata Mahajan
B.Tech, Computer Science Engineering
Pimpri Chinchwad College of Engineering, Pune, India
akshata.mahajan22@pccoepune.org

Sayee Patil
B.Tech, Computer Science Engineering
Pimpri Chinchwad College of Engineering, Pune, India
sayee.patil22@pccoepune.org

Rutuja Kalekar
B.Tech, Computer Science Engineering
Pimpri Chinchwad College of Engineering, Pune, India
rutuja.kalekar22@pccoepune.org

Pranjali Kotalwar
B.Tech, Computer Science Engineering
Pimpri Chinchwad College of Engineering, Pune, India
pranjali.kotalwar22@pccoepune.org

Abstract
This study explores the integration of behavioral metrics into traditional student evaluation frameworks to gain deeper insights into test-taking strategies and performance. Metrics such as total time spent on tests, time efficiency, correct- to-incorrect answer ratios, and question revisit proportions are quantified using well-defined formulas. The aim is to complement traditional grading systems by providing educators with actionable insights into time management, decision-making, and confidence levels. This paper highlights the methodology for metric computation, discusses findings using simulated datasets, and illustrates how these metrics can enhance teaching strategies and support personalized feedback.

Index Terms
Student Performance Analysis, Online Assessment, Behavioral Metrics, Holistic Evaluation, Learning Analytics.

1. Introduction

In the evolving landscape of education, assessment has remained a cornerstone for measuring student understanding and academic progress. Traditionally, evaluations rely predominantly on correctness-based grading systems, where the accuracy of answers serves as the primary metric for gauging performance. While effective in assessing knowledge acquisition, these conventional approaches often fail to capture the nuanced behavioral patterns exhibited by students during tests. Such patterns include time allocation across questions, the frequency of revisiting previously answered items, and the tendency to revise answers. These behavioral elements can provide valuable insights into cognitive processes, time management skills, decision-making strategies, and levels of confidence, yet they are time management or surface-level engagement with the material. Recognizing and addressing these patterns can empower educators to adopt more targeted teaching interventions and feedback strategies tailored to individual needs.

This study introduces a holistic evaluation system by integrating behavioral metrics into traditional grading method- ologies. Metrics such as average time per question, time efficiency, correct-to-incorrect answer ratios, revisit proportions, and answer changes are analyzed to uncover hidden patterns in student performance. By correlating these metrics with traditional performance indicators, the research aims to bridge the gap between correctness- based evaluations and behavioral insights. The findings hold the potential to redefine educational assessments, making them more reflective of a student's overall learning journey rather than just their ability to recall or reproduce information under test conditions. The paper presents a comprehensive framework for incorporating these behavioral metrics into assessment systems. It outlines the methodology for data collection and computation, discusses the significance of the derived metrics, and highlights their implications using simulated remain largely unexplored in typical evaluation frameworks. Research has shown that understanding these behavioral dimensions is crucial for addressing broader educational goals, such as fostering critical thinking and enhancing learning experiences. For instance, students who frequently revisit questions or change answers might be struggling with confidence or experiencing decision fatigue, while those who rush through questions may demonstrate poor student data. By doing so, this study seeks to provide educators with a deeper understanding of the learning process, enabling them to tailor teaching strategies, improve test design, and offer constructive feedback that promotes long term academic and personal growth. Ultimately, this approach aims to foster a more comprehensive evaluation paradigm that prioritizes not just what students learn, but also how they approach learning challenges.

2. Methodology

A. Form Design and Logging Mechanism

To gather meaningful behavioral data for performance evaluation, a simulated online assessment platform was designed to mimic real-world digital test environments. The

form was developed with integrated logging features to track not only student responses but also test-taking behaviors such as time spent, navigation patterns, and decision changes.

Simulated Test Interface: The test environment resembled a standard multiple choice digital exam, where students could navigate between questions and modify their answers. Each student record was captured with the assumption that they interacted with the test under conditions similar to online assessments used in platforms like Moodle or Google Forms with scripting.

Behavioral Tracking Features: The form recorded detailed behavioral metadata using embedded logging mechanisms:

- Time Tracking: Captured time spent on each individual question.
- Revisit Monitoring: Logged how many times a student navigated back to a previously viewed question.
- Answer Change Detection: Recorded when students changed their selected answer after initial submission. These logs were collected in addition to the final responses and correctness to enable a process-oriented analysis, not just an outcome based evaluation.

Example Data Captured: Each virtual student record in the simulated dataset included:

- Student ID (unique identifier)
- Question wise timestamps (start and end times per question)
- Initial answer and final answer for each question
- Revisit status (e.g., visited once or multiple times)
- Answer change count
- Total time spent on test
- Number of questions attempted
- Number of correct and incorrect responses

Purpose of Behavioral Tracking: The inclusion of behavioral logs allows for:

- Identifying hesitation patterns through revisits and answer changes
- Assessing confidence levels and time management strategies
- Differentiating between rushed guessing and thoughtful answering
- Providing educators with insights that go beyond correct or incorrect grading

This data collection approach sets the foundation for computing behavioral metrics such as revisit proportion, time efficiency, and answer change frequency, which are discussed in the following sections.

B. Data Collection

The data for this research was simulated using synthetic datasets modeled on realistic test-taking behaviors. A sample of 100 virtual student records was generated, mimicking test attempts on an online assessment platform. Each record included:

- Total time spent on the test: Captures pacing at the overall level.
- Time spent on each question: Highlights question-specific time management.
- Revisit counts: Tracks how often students returned to previous questions.
- Answer changes: Logs the frequency of revisions to previously selected answers.
- Correct and incorrect answers: Tracks overall performance.

C. **Metrics for Analysis**

1) Performance Metrics:
   - Accuracy (%): Ratio of correct answers to total questions.
   - Correct to Incorrect Ratio: Indicates balance in performance.

2) Behavioral Metrics:
   - Answer Changes (%): Frequency of changing answers.
   - Revisit Count: Measures hesitation or review tendency.

3) Time Management Metrics:
   - Average Time per Question
   - Time Efficiency: Number of correct answers relative to total time.

4) Computed Metrics:
   - Average Time per Question: Highlights time management efficiency.

$$\textbf{\textit{Average Time per Question}} = \frac{Total\ Time\ Spent\ on\ Test}{Number\ of\ Quesetions\ Answered}$$

- Accuracy: Measures test performance.

$$Accuracy\ (\%) = \left(\frac{Number\ of\ Correct\ Answers}{Total\ Number\ of\ Questions}\right) x100$$

- Correct – to – Incorrect Ratio: Reflects preparation quality.

$$Correct - to - Incorrect\ Ratio = \frac{Number\ of\ Correct\ Answers}{Number\ of\ Incorrect\ Answers}$$

- Proportion of Answer Changes: Indicates indecision or lack of confidence.

$$Proportaion\ of\ Answer\ Changes\ (\%) = \left(\frac{Number\ of\ Questions\ with\ Answer\ Changes}{Total\ Numebr\ of\ Questions}\right) x100$$

- Revisit Proportion Purpose: Identifies hesitation or second-guessing tendencies

$$Revisit\ Proportion\ (\%) = \left(\frac{Number\ of\ Visited\ Questions}{Total\ Number\ of\ Questions}\right) x100$$

- Tiem Efficiency: Evaluates how effectively students manage time to answer questions correctly.

$$Time\ Efficiency = \frac{Number\ of\ Correct\ Answers}{Total\ Time\ Spent\ (minutes)}$$

- Overall Performance Rating (Weighted Average): Accounts for questions-specific importance.

$$Overall\ Performance = \frac{\sum_{i=1}^{n}(Score\ on\ Question\ i\ x\ Weight\ of\ Question\ i)}{\sum_{i=1}^{n}(Weight\ of\ Quesion\ i)}$$

D. Data Analysis Techniques

A combination of descriptive statistics and correlation analyses was employed:

- Descriptive Statistics: Used to summarize data trends (mean, median, and standard deviation).
- Correlation Analysis: Explored relationships between behavioral metrics (e.g., revisit frequency) and performance outcomes (e.g., accuracy).

Graphs and tables were generated to visualize these relationships, providing concrete evidence of metric significance.

Table I: Key Metric Statistics

Metric	Mean	Median	Std Dev
Accuracy (%)	75.2	74.5	8.3
Correct-to-Incorrect Ratio	1.8	1.7	0.4
Avg Time per Question (sec)	42.3	40.1	9.2
Revisit Frequency (%)	12.7	10.2	6.8
Answer Change Frequency (%)	8.3	6.9	4.5

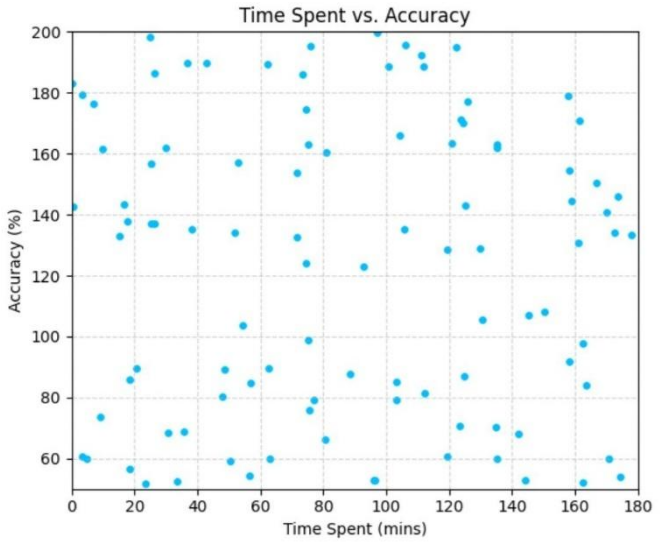

Fig. 1. Scatter plot showing correlation between Time Spent and Accuracy.

3. Results and Analysis

A. Descriptive Statistics.

Table I summarizes the descriptive statistics for the key metrics:

B. Correlation Findings

The key findings from the correlation analysis are:

- Time Spent vs. Accuracy: Moderate positive correlation ($\rho = 0.45$), suggesting that students who spent more time per question performed better, up to an optimal threshold.
- Revisit Frequency vs. Accuracy: Weak negative correlation ($\rho = -0.18$), indicating that frequent revisits were slightly associated with lower scores, possibly due to test anxiety or overthinking.
- Answer Changes vs. Final Score: Students who frequently revised answers often scored lower, suggesting overcorrection tendencies.

C. Key Insights

- Time Efficiency and Accuracy: Students with balanced time allocation demonstrated higher accuracy, while those with extreme values (too fast or too slow) underperformed.
- Behavioral Trends: High revisit proportions were associated with lower scores, suggesting hesitation.

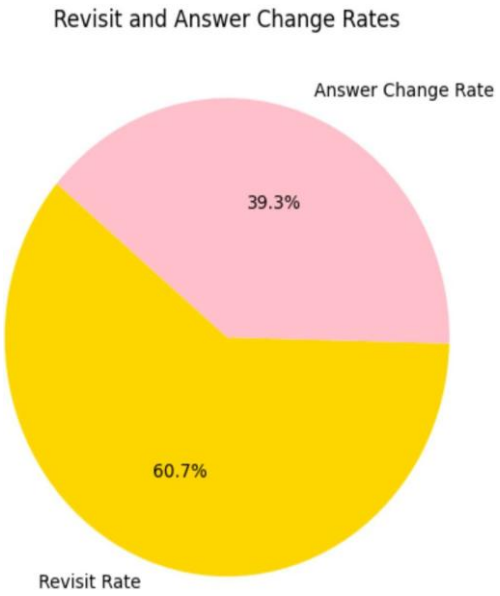

Fig. 2. Distribution of Answer Change Frequency among students.

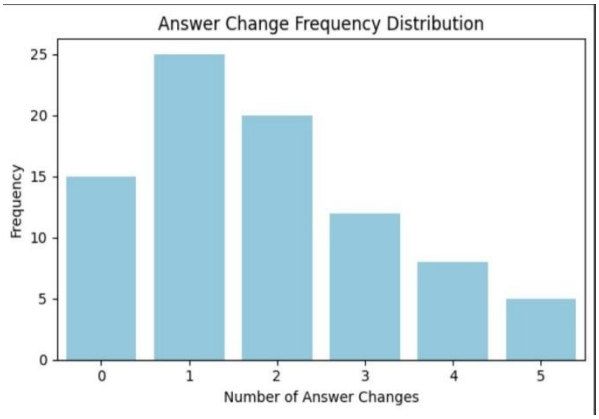

Fig. 3. Proportion of Revisit Rate vs. Answer Change Rate.

Frequent answer changes correlated negatively with performance, indicating overcorrection.

Correlation Analysis:
Moderate positive correlation between time spent per question and accuracy ($\rho = 0.42$).
Negative correlation between answer changes and final scores ($\rho = -0.21$).

4. Discussion
The analysis reveals significant insights into the impact of test-taking behaviors on student performance:

Time Management: Students who demonstrated balanced pacing performed better overall, while those who spent too much or too little time per question often underperformed.

Confidence and Decision Making: Frequent revisits and answer changes correlated with lower scores, indicating potential test anxiety or lack of confidence.

Holistic View: Behavioral metrics provide actionable insights that traditional score-based evaluations miss.

5. Insights
Practical Insights for Educators and Learners: The integration of behavioral metrics into student assessment systems offers actionable benefits for both educators and learners, bridging the gap between raw scores and learning strategies.

For Educators: Behavioral data provides instructors with granular insight into how students approach assessments, not just how many answers they get right or wrong. This enables:

- Personalized Feedback: If a student revisits many questions or frequently changes answers, teachers can infer possible test anxiety or conceptual uncertainty and ad- dress it in follow-up sessions.
- Improved Test Design: Questions with disproportionately high revisit or answer change rates might be too ambiguous or poorly worded, allowing educators to refine assessment quality.
- Differentiated Instruction: Teachers can group students based on patterns like rushing, overthinking, or indecisiveness and create tailored interventions for each cate- gory.

For Students: Students also benefit significantly from understanding their behavioral performance metrics. Instead of relying solely on final scores, they can:

- Develop Self-Awareness: Recognize whether they spend too much time on certain questions or change answers unnecessarily, helping them adjust strategies in future tests.
- Improve Confidence and Pacing: High revisit or answer change frequencies may indicate a lack of confidence; once identified, students can consciously work on their decision-making and time allocation.
- Shift from Score-Only Mindset: By seeing test-taking as a skill, students start valuing behavioral control alongside subject mastery.

Behavior-Aware vs. Score-Based Evaluation
While traditional systems focus purely on outcomes (i.e., marks), behavior-aware evaluation highlights the process behind those outcomes. This dual insight model helps uncover learning difficulties not visible through scores alone. It encourages a more supportive and diagnostic approach to education, where the aim is not just performance measurement but performance improvement.

6. Conclusion
This study presents a framework for holistic student assessment by integrating behavioral metrics, such as time

spent per question, correct-to-incorrect ratios, and time efficiency,

into traditional evaluation systems. These metrics provide deeper insights into cognitive strategies, decision-making, and confidence, enabling educators to offer more personalized feedback and address specific learning gaps.

The findings highlight the potential of behavioral metrics to improve educational outcomes by identifying areas where students excel or need support. This approach benefits individual learners through targeted interventions and helps institutions refine test design and teaching strategies.

Future research will validate this framework with real-world datasets, explore its longitudinal impact, and investigate its integration into adaptive learning platforms. By recognizing test-taking behaviors, this framework promotes more comprehensive evaluations, fostering personalized learning and long- term academic growth.

7. References

[1] Hassan, Muhammad Awais, et al. ”An adaptive feedback system to improve student performance based on collaborative behavior.” IEEE Access, vol. 7, 2019, pp. 107171–107178.

[2] Serrano-Aguilera, Juan Jose, et al. ”Using peer review for student performance enhancement: Experiences in a multidisciplinary higher education setting.” Education Sciences, vol. 11, no. 2, 2021, p. 71.

[3] Barik, Lalbihari, Omar Barukab, and A. Ahmed. ”Employing artificial intelligence techniques for student performance evaluation and teaching strategy enrichment: An innovative approach.” International Journal of Advanced and Applied Sciences, vol. 7, no. 11, 2020, pp. 10–24.

[4] Singh, Amika S., et al. ”Effects of physical activity interventions on cognitive and academic performance in children and adolescents: a novel combination of a systematic review and recommendations from an expert panel.” British Journal of Sports Medicine, vol. 53, no. 10, 2019, pp. 640–647.

[5] Thaher, Thaer, et al. ”An enhanced evolutionary student performance prediction model using whale optimization algorithm boosted with sine-cosine mechanism.” Applied Sciences, vol. 11, no. 21, 2021, p. 10237.

[6] Almufarreh, Ahmad. ”Exploring the potential of mixed reality in enhancing student learning experience and academic performance: an empirical study.” Systems, vol. 11, no. 6, 2023, p. 292.

[7] Shawon, Reza E. Rabbi, Mohammed Nazmul Islam Miah, and Md Zahidul Islam. ”Enhancing US Education Systems with AI: Personalized Learning and Academic Performance Prediction.” International Journal of Advanced Engineering Technologies and Innovations, vol. 1, no. 1, 2023, pp. 518–540.

[8] Bailey, Regan L., et al. ”Redesigning an undergraduate nutrition course through active learning and team-based projects enhances student performance.” Current Developments in Nutrition, vol. 4, no. 4, 2020, p. nzaa039.

[9] Liu, Yutong, et al. ”Predicting student performance using clickstream data and machine learning.” Education Sciences, vol. 13, no. 1, 2022, p. 17.

[10] Ni, Lin, et al. ”Enhancing student performance prediction on learner-sourced questions with SGNN-LLM synergy.” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 21, 2024.

[11] Gonza´lez-Calatayud, V´ıctor, Paz Prendes-Espinosa, and Rosabel Roig-Vila. ”Artificial intelligence for student assessment: A systematic review.” Applied Sciences, vol. 11, no. 12, 2021, p. 5467.

[12] Prigoff, Jake, Madeleine Hunter, and Roman Nowygrod. ”Medical student assessment in the time of COVID-19.” Journal of Surgical Education, vol. 78, no. 2, 2021, pp. 370–374.

[13] Jacob, Brian, and Jesse Rothstein. ”The measurement of student ability in modern assessment systems.” Journal of Economic Perspectives, vol. 30, no. 3, 2016, pp. 85–108.

[14] Hirsch, Adina C., and Harish S. Parihar. ”A capstone course with a comprehensive and integrated review of the pharmacy curriculum and student assessment as a preparation for advanced pharmacy practice experiences.” American Journal of Pharmaceutical Education, vol. 78, no. 10, 2014, p. 192.

[15] Kartowagiran, Badrun, et al. Can student assessment sheets replace observation sheets? Yogyakarta State University, 2019.